International Journal of Mechanical Engineering and Technology (IJMET)
Volume 10, Issue 1, January 2019, pp. 344–351, Article ID: IJMET_10_01_035
Available online at http://www.iaeme.com/ijmet/issues.asp?JType=IJMET&VType=10&IType=1
ISSN Print: 0976-6340 and ISSN Online: 0976-6359

CLASSIFICATION OF THE LATEST HANDPHONE PRODUCTS IN THE TOKO REJEKI CELULAR IN MERAUKE DISTRICT

Gerzon Jokomen Maulany and Nasra Pratama Putra
Department of Information Systems, Faculty of Engineering,
Universitas Musamus, Merauke, Indonesia

ABSTRACT

Toko Rejeki Celular is a shop that sells a variety of telecommunications electronic equipment, one of which is a handphone. The store manager is required to be able to make the right decisions in determining the sales strategy. In order to do this, further analysis is needed regarding data from the sale of mobile phones and the needs of customers. The purpose of this study was to apply data mining techniques to the Rejeki Celular Shop in Merauke Regency. The results of the study are expected to provide information in the form of classifications of sales of mobile phones that are most popular with customers and are less popular (best sales and normal sales). The data mining method used is the decision tree method, where the algorithm used is the C45 algorithm. As for the attributes are the type of mobile phone, price range, battery size and screen size. The data sample used is 21 data which is the sales data for mobile phones for 1 month. The results of this study are in the form of a system built using the PHP programming language and MySQL database. The highest factor affecting the purchase of mobile phones at Toko Rejeki Celular is the type of mobile attribute with the highest Gain, which is equal to 0.21687. The next factor is the price range attribute. As for the battery capacity factor and screen size it has no effect in producing a decision tree.

Keywords: Data Mining, Decision Tree, Algorithm C45, Handphone

Cite this Article: Gerzon Jokomen Maulany and Nasra Pratama Putra, Attitude Quadrotor Control System with Optimization of PID Parameters Based On Fast Genetic Algorithm, International Journal of Mechanical Engineering and Technology, 10(1), 2019, pp. 344–351.
http://www.iaeme.com/ijmet/issues.asp?JType=IJMET&VType=10&IType=1

INTRODUCTION

Toko Rejeki Celular was founded in 2005, where it served the sale of various telecommunications electronic equipment such as mobile phones, laptops, netbooks, pagers, and various accessories for various such equipment. Besides serving telecommunication

electronic goods sales, Toko Rejeki Celular also provides service for damage services on mobile phones, and sells various applications, themes, games used for computers and mobile phones. The types of cellphones sold include the Nokia brand, Blackberry, Samsung, Advan, Cross, Htc, Mito, and many more in accordance with the development of the type of cellphone each year. The increase in service continues to be carried out by the manager. Currently, Toko Rejeki Celular has used the application in the transaction process. Although it has been supported by technological and information capabilities, store managers still find it difficult to obtain strategic information. One of the most needed information is about the best-selling products, especially on mobile phones.

The rapid development of technology and information has triggered the creation of smart innovations in business. This fact on the one hand has helped companies to increase company revenue by finding new business opportunities and creating competitive advantages. This advancement in technology and information is often known as business intelligence or business intelligence. Data Mining Technology is one way to extract useful information from sales company data warehouses. Managers can use the data warehouse they already have to find useful information to help draw conclusions. The problem is the large amount of data that gives rise to a new branch of science to overcome the problem or what is known as data mining. At a further stage, data mining techniques are expected to be able to explore a variety of new knowledge that was previously unknown.

Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases. One technique that exists in data mining is classification (Efraim et al., 2005). Previous research on data mining utilization in business included: classification of bank customer data for credit decision making (Rani, 2016), prediction of number of motorcycle sales (Azwanti, 2018), sales partner recommendation acceptance (Arifinand Firianah, 2018), strategic information on batik sales (Nugrohoand Alirsyadi, 2015), selling best-selling products distro (Winata, 2017), classification of behavior of purchasing patterns (Susanto, et al., 2015) and classification of customer loyalty to product brands (Ariwibowo, 2013).

The importance of this research is to try to apply data mining techniques using the C4.5 algorithm decision tree method. The C45 algorithm is applied to calculate the classification of the Celular Fortune Shop in Merauke Regency. The results of the analysis are expected to provide information in the form of classifications of sales of mobile phones that are most popular with customers and less popular (best sales and normal sales). Based on these results, the shop owner Rejeki Celular as the manager can analyze the procurement of stock mobile phones that follow the trends and preferences of their customers.

METHODOLOGY

The method used in this study is the C.45 Algorithm. To be able to use the C45 Algorithm, five KDD steps are taken (Knowledge Discovery in Databases). These steps are data collection, selection, preprocessing, transformation, data mining, interpretation and evaluation.

1. Data collection

   The data needed in this study is the sales data of mobile phones from the RejekiCelular Shop. The data will be used as training data. Training data is the data needed to get a pattern generated from sales data that was previously available.

2. Data selection
   Data selection is the process used to select only a portion of the data needed. The main purpose of the data selection process is to create a target data set, selection of data sets, or focus on a subset of variables or sample data, where discovery will be carried out next (Hirianaand Rasyidan,2017). The variables needed in this study are:
   - Variable X1 handphone_type
   - Variable X2 level_price
   - Variable X3 battery_capacity
   - Variable X2 screen_size
   - Variable Y1 decision
3. Data transformation
   Data transformation is the process of transforming or converting variables into appropriate forms, namely attributes and values. The variables used are in Table 1.

Table 1. Attribute and Value Category

Attribute	Value	Type
handphone_type	Gaming Normal Photography	Discrete
level_price	<1500000 1500000-2500000 >2500000	Discrete
battery_capacity	>4000 mah 3500-4000 mah <3500 mah	Discrete
screen_size	<5 inch 5-6 inch >6 inch	Discrete
decision	Best Selling Normal Selling	Discrete

4. Data mining

Data mining is an integral part of knowledge discovery in databases which is a process in the following order (Tyas et al., 2010):
- Data cleaning (to eliminate noise and data inconsistencies)
- Data integration (some data sources will be combined)
- Data selection (only data that can be used for analysis will be taken from the database)
- Data transformation (data will be transformed into more structured forms to simplify the data mining process)
- Data mining (the main process of data mining where data mining techniques are applied)
- Pattern evaluation
- Knowledge presentation (where visualization and results representation is given to users)

The main purpose of applying data mining is to predict (prediction) and description (description). Classification is the process of finding a model (or function) that will classify data classes so that they can predict classes of unknown objects. In this study using the C45 algorithm.

In general, the steps taken by the C45 algorithm to build a decision tree are as follows:

Input: sample training, label training, attribute.

1. Make a root node for the tree that is made
2. If all samples are positive, stop with a tree with one root node, give a sign (+)
3. If all samples are negative, stop with a tree with one root node, give a sign (-)
4. If the attribute is empty, stop with a tree with a root node, with the label corresponding to the most value that is on the training label
5. For others, Start
   a. A ---- attributes that classify samples with the best results (based on gain ratio)
   b. Decision attribute for root node ---- A
   c. For each value, vi, which is possible for A
      (1) Add a branch under the root associated with A = vi
      (2) Determine the sample SRI as a subset of samples that have a value of vi for attribute A
      (3) If the sample Svi is empty
         i. Under the branch add a leaf node with the label = the most value on the training label
         ii. The others add new branches under branches which are now C 4.5 (sample training, training labels, attributes - [A].
   d. Stop

Output: Decision Tree.

To choose an attribute as root, it is based on the highest gain value of the existing attributes. Information gain is one of the attribute selection measures used to select test attributes for each node in the tree. Attributes with the highest gain information are selected as test attributes of a node. Gain (S, A) is the acquisition of information from attribute A relative to the output of data S. The acquisition of information obtained from output data or dependent variable S grouped by attribute A, denoted by gain (S, A). To calculate gain, use equation (1).

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{S} Entropy(S_i) \quad (1)$$

Where:

S = Set of cases

A = Attribute

n = Number of partition attributes A

$|Si|$ = Number of cases on the i partition

$|S|$ = Number of cases in S

While entropy is a measure of information theory that can know the characteristics of impurity and homogeneity of the data set. From the Entropy value, then the value of information gain (IG) is calculated for each attribute. Entropy (S) is the number of bits expected to be able to extract a class (+ or -) from a number of random data in a sample space S. Entropy can be said as a bit requirement to declare a class. The smaller the Entropy value, the more Entropy is used to extract a class. Entropy is used to measure the authenticity of the S or the processing system. To calculate the entropy value used equation (2).

$$Entropy(S) = \sum_{i=1}^{n} -P_i log_a P_i \quad (2)$$

Where:

S = Set of cases
n = Number of partitions S
a = Feature
Pi = Proportion of |Si| against S

5. Implementation

At this stage the results of the analysis are implemented in the form of applications that can provide a determination of the classification of sales of the best-selling mobile products.

In addition, other useful methods were provided (Istanto and Manggau, 2018; Lamalewa and Maulany, 2018; Latuheru and Sahupala, 2018; Waremra and Bahri, 2018).

RESULTS

Data analysis

The data used is an example of handphone sales data at the Fortune Celular Shop in one month (30 days). It is known that several factors that determine the purchase of mobile phones are the type of mobile phone, price range, battery capacity and screen size. The pre-process results carried out on the sales data can be seen in Table 2.

Table 2. Pra-Process Value

Barcode	Handphone Type	Level Price	Battery Capacity	Screen Size	Decision
BRG_1021	Normal	1500000-2500000	>4000 mah	<5 inch	Best Selling
BRG_1022	Normal	1500000-2500000	>4000 mah	5-6 inch	Best Selling
BRG_1023	Photography	>2500000	>4000 mah	>6 inch	Best Selling
BRG_1024	Normal	<1500000	>4000 mah	5-6 inch	Best Selling
BRG_1025	Normal	<1500000	<3500 mah	<5 inch	Best Selling
BRG_1026	Gaming	<1500000	<3500 mah	<5 inch	Best Selling
BRG_1027	Normal	<1500000	>4000 mah	<5 inch	Normal Selling
BRG_1028	Normal	>2500000	>4000 mah	<5 inch	Best Selling
BRG_1029	Normal	>2500000	3500-4000 mah	>6 inch	Best Selling
BRG_1030	Normal	>2500000	<3500 mah	<5 inch	Best Selling
BRG_1031	Normal	1500000-2500000	<3500 mah	<5 inch	Normal Selling
BRG_1032	Normal	1500000-2500000	<3500 mah	5-6 inch	Normal Selling
BRG_1033	Photography	<1500000	>4000 mah	<5 inch	Normal Selling
BRG_1034	Photography	<1500000	>4000 mah	5-6 inch	Normal Selling
BRG_1035	Photography	>2500000	<3500 mah	<5 inch	Normal Selling
BRG_1036	Photography	>2500000	<3500 mah	5-6 inch	Normal Selling
BRG_1037	Normal	>2500000	>4000 mah	<5 inch	Best Selling
BRG_1038	Normal	>2500000	>4000 mah	5-6 inch	Best Selling
BRG_1039	Gaming	1500000-2500000	>4000 mah	>6 inch	Normal Selling
BRG_1040	Gaming	1500000-2500000	3500-4000 mah	5-6 inch	Normal Selling
BRG_1041	Gaming	1500000-2500000	<3500 mah	<5 inch	Normal Selling

Entropy and Gain Calculation

Based on equations (1) and (2), the initial calculation is performed to determine the root of the decision tree. The results of the calculations can be seen in Table 3.

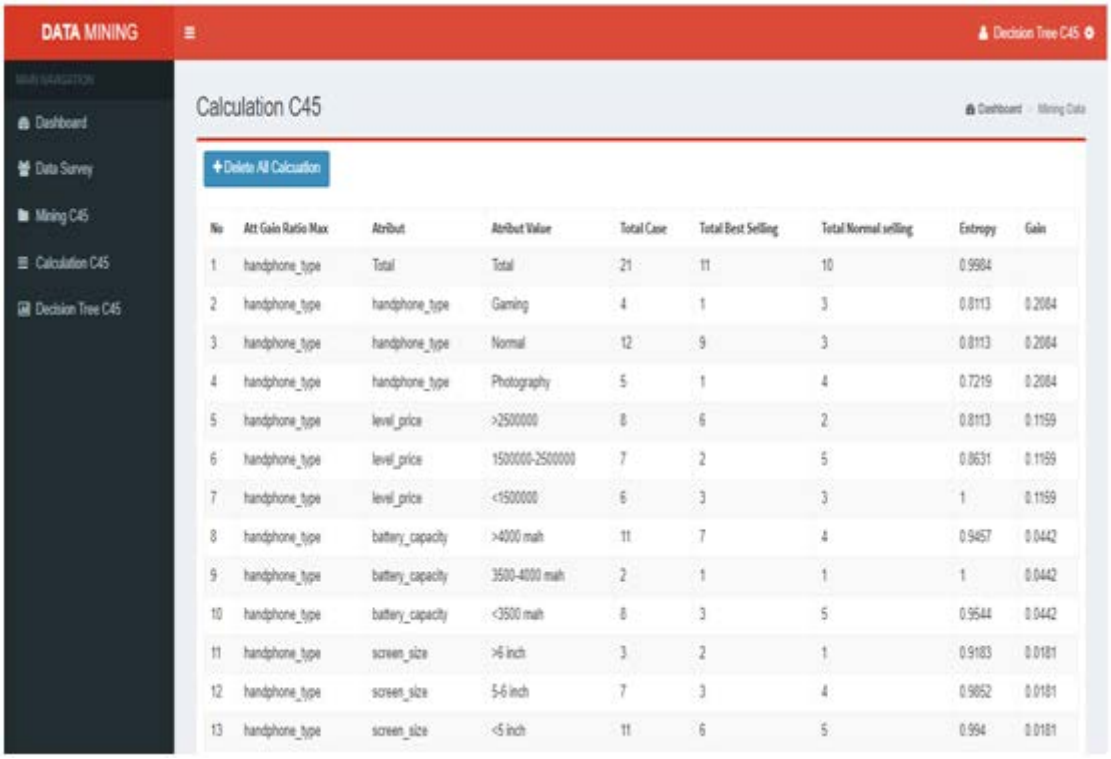
Table 3. Root Node Calculation Results

Node			Total Case	Best Selling	Normal Selling	Entropy	Gain
1	Total		21	10	11	0.99836	
	Handphone Type						0.21687
		Gaming	4	3	1	0.81128	
		Normal	12	3	9	0.81128	
		Photography	5	4	1	0.72193	
	Level Price						0.11588
		>2500000	8	2	6	0.81128	
		1500000-2500000	7	5	2	0.86312	
		<1500000	6	3	3	1	
	Battery Capacity						0.04419
		>4000 mah	11	4	7	0.94566	
		3500-4000 mah	2	1	1	1	
		<3500 mah	8	5	3	0.95443	
	Screen Size						0.01809
		>6 inch	3	1	2	0.9183	
		5-6 inch	7	4	3	0.98523	
		<5 inch	11	5	6	0.99403	

From the results in table 3 it can be seen that the attribute with the highest Gain is Mobile Type, which is equal to 0.21687. So that it can be said to attribute the Mobile Type as the root node. There are three more attributes that must be calculated using the C45 algorithm, namely Price Level, Battery Capacity and Screen Size.

Implementation

System implementation is done by calculating the entropy and gain values using the PHP programming language and MySql database. The system is built in the form of a website because the php programming language is an open source programming language. The results of the whole calculation in Table 3 then proceed to the program as in Figure 1. Whereas the results of decision tree formation can be seen in Figure 2.



No	Att Gain Ratio Max	Atribut	Atribut Value	Total Case	Total Best Selling	Total Normal selling	Entropy	Gain
1	handphone_type	Total	Total	21	11	10	0.9984	
2	handphone_type	handphone_type	Gaming	4	1	3	0.8113	0.2084
3	handphone_type	handphone_type	Normal	12	9	3	0.8113	0.2084
4	handphone_type	handphone_type	Photography	5	1	4	0.7219	0.2084
5	handphone_type	level_price	>2500000	8	6	2	0.8113	0.1159
6	handphone_type	level_price	1500000-2500000	7	2	5	0.8631	0.1159
7	handphone_type	level_price	<1500000	6	3	3	1	0.1159
8	handphone_type	battery_capacity	>4000 mah	11	7	4	0.9457	0.0442
9	handphone_type	battery_capacity	3500-4000 mah	2	1	1	1	0.0442
10	handphone_type	battery_capacity	<3500 mah	8	3	5	0.9544	0.0442
11	handphone_type	screen_size	>6 inch	3	2	1	0.9183	0.0181
12	handphone_type	screen_size	5-6 inch	7	3	4	0.9852	0.0181
13	handphone_type	screen_size	<5 inch	11	6	5	0.994	0.0181

Figure 1. Calculation C45

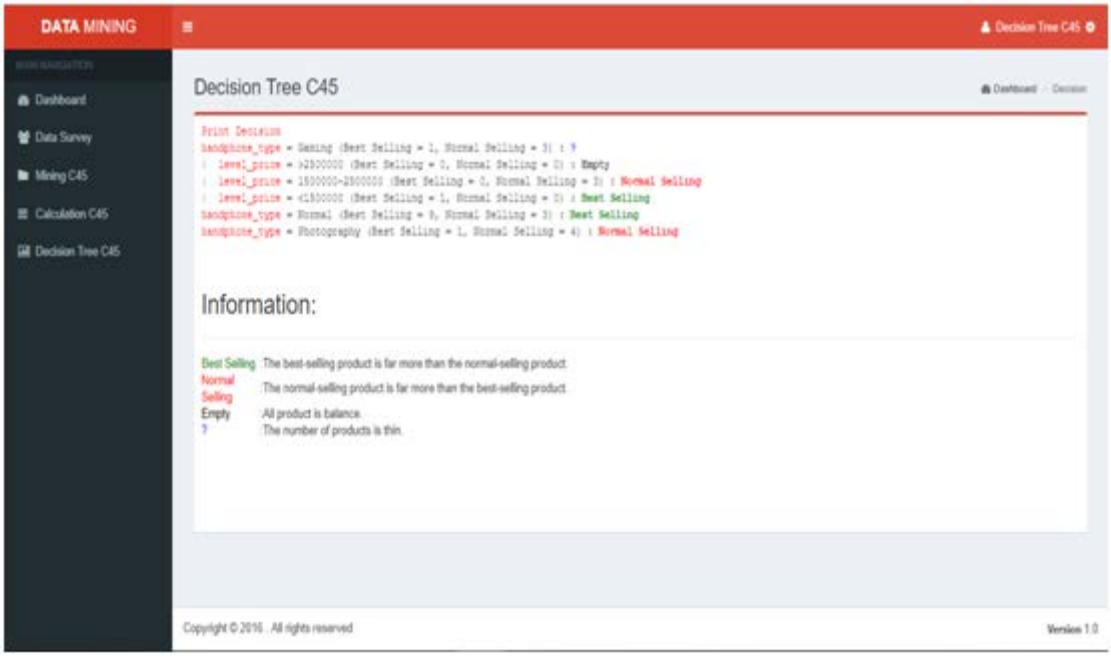


Figure 2. Results of the Decision Tree

CONCLUSIONS

- Based on the results of the research conducted at the Fortune Celular Shop, it can be concluded that the use of the Data Mining method especially the C4.5 Algorithm will be very useful for managers in the decision making process for stocking mobile phones.
- The highest factor affecting the purchase of mobile phones at the Rejeki Celular Shop is the handphone type attribute.

- The second factor that influences mobile phone purchases at the Fortune Celular Shop is the price range attribute.
- Battery capacity and screen size factors have no effect in producing decision trees.

REFERENCES

[1] Arifin, M.F. dan Firianah, Devi. 2018. "Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada" IncomTech, Jurnal Telekomunikasi dan Komputer, vol.8, no.2 Halaman 87-102, ISSN : 2085-4811.

[2] Ariwibowo, A.S. 2013. "Metode Data Mining Untuk Klasifikasi Kesetiaan Pelanggan Terhadap Merek Produk" Seminar Nasional Sistem Informasi Indonesia, Halaman 532-540, 2-4 Desember 2013.

[3] Azwanti, Nurul. 2018. "Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada PT. Capella Dinamik Nusantara Cabang Muka Kuning", Jurnal Ilmiah Ilmu Komputer, Informatika Mulawarman Vol 13 No. 1 Halaman 33-38, E-ISSN : 2597-4963 P-ISSN : 1858-4853, Februari 2018.

[4] Efraim Turban, Jay E. Aronson, Ting Peng Liang, 2005. *Decision Support System and Intelligent Systems Edisi 7 Jilid 1*, Andi Yogyakarta.

[5] Hiriana, Nadiya dan Rasyidan, Muhammad. 2017. "Penerapan Metode Decision Tree Algoritma C4.5 untuk Seleksi Calon Penerima Beasiswa Tingkat Universitas" Al Ulum Sains dan Teknologi, Vol. 3 No. 1 Halaman 9-13, November 2017.

[6] Istanto, T and Manggau, FX. 2018. Analysis of the Power of Wifi Signals on the Informatics Engineering Laboratory of Musamus University Using Insidder, International Journal of Mechanical Engineering and Technology 9(13), pp.266–272.

[7] Lamalewa, L. and Maulany, GJ. 2018. Application of Case Based Reasoning and Nearest Neighbor Algorithm for Positioning Football Players, International Journal of Mechanical Engineering and Technology 9(13), pp.258–265.

[8] Latuheru, RD and Sahupala, P. 2018. Design of Methane Gas Capture Installation from Garbage Waste, International Journal of Mechanical Engineering and Technology, 9(10), pp. 279–287.

[9] Nugroho, Y.S. dan Alirsyadi, F.Y. 2015, "Implementasi Data Mining Sebagai Informasi Strategis Penjualan Batik (Studi Kasus Batik Mahkota Laweyan)" Prosiding SNATIF ke-2, Halaman 161-168, ISBN: 978-602-1180-21-1, Universitas Muria Kudus, 2015

[10] Rani, L.N. 2016, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit" Jurnal InovtekPolbeng – Seri Informatika Vol. 1 No. 2 Halaman 126-132, ISSN : 2527-9866, November 2016.

[11] Susanto, et al., 2015. "Penerapan Data Mining Classification Untuk Prediksi Perilaku Pola Pembelian Terhadap Waktu Transaksi Menggunakan Metode Naïve Bayes" Konferensi Nasional Sistem dan Informatika STMIK STIKOM, Halaman 313-318, Bali, 9-10 Oktober 2015.

[12] Tyas, A.F. et al. 2010, "Klasifikasi Data Dengan Menggunakan Algoritma C4.5 dan TAN" Tugas Akhir Program Studi S1 Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom, Halaman 15-61, 2010.

[13] Waremra, RS and Bahri, S. 2018. Identification of Light Spectrum, Bias Index and Wavelength in Hydrogen Lights and Helium Lights Using a Spectrometer, International Journal of Mechanical Engineering and Technology 9(10), pp. 72–76.

[14] Winata, A.D. 2017, "Analisis dan Prediksi Penjualan Produk Terlaris Distro Root Shoes Dengan Aplikasi Android" Artikel Publikasi Universitas Muhammadiyah Surakarta.